

Soil Organic Carbon Data Rescue and Harmonization Repository (SOC-DRaHR)

Kathe Todd-Brown (ISCN, PNNL)

6 February 2018

International Soil Carbon Network

Leverage large scale synthesis studies to inform soil carbon understanding by:

- 1) Identify **big data questions**
- 2) Develop a **common vocabulary** around soil measurements
- 3) Harmonize **soil data** from surveys and experiments (both field and lab)

ISCN both a network and data product (ISCN3).

Problem: Lots of data in repository but hard to use.

Solution: A community open source software project to identify datasets of interest and create a collection of R scripts to harmonize that data.

SOC-DRaHR

Traditional meta-analysis pipeline

1. Create template
2. Give template to data providers
3. Merge template
4. Analyze resulting database

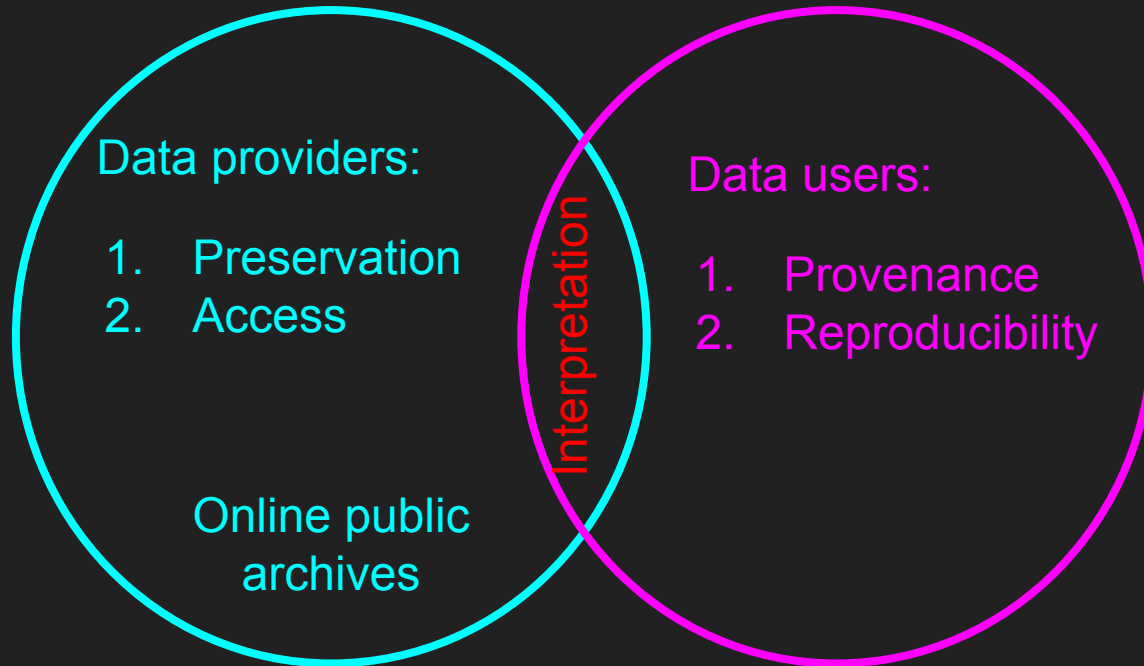
Traditional meta-analysis pipeline

1. Create template
2. Give template to data providers
3. Merge template
4. Analyze resulting database

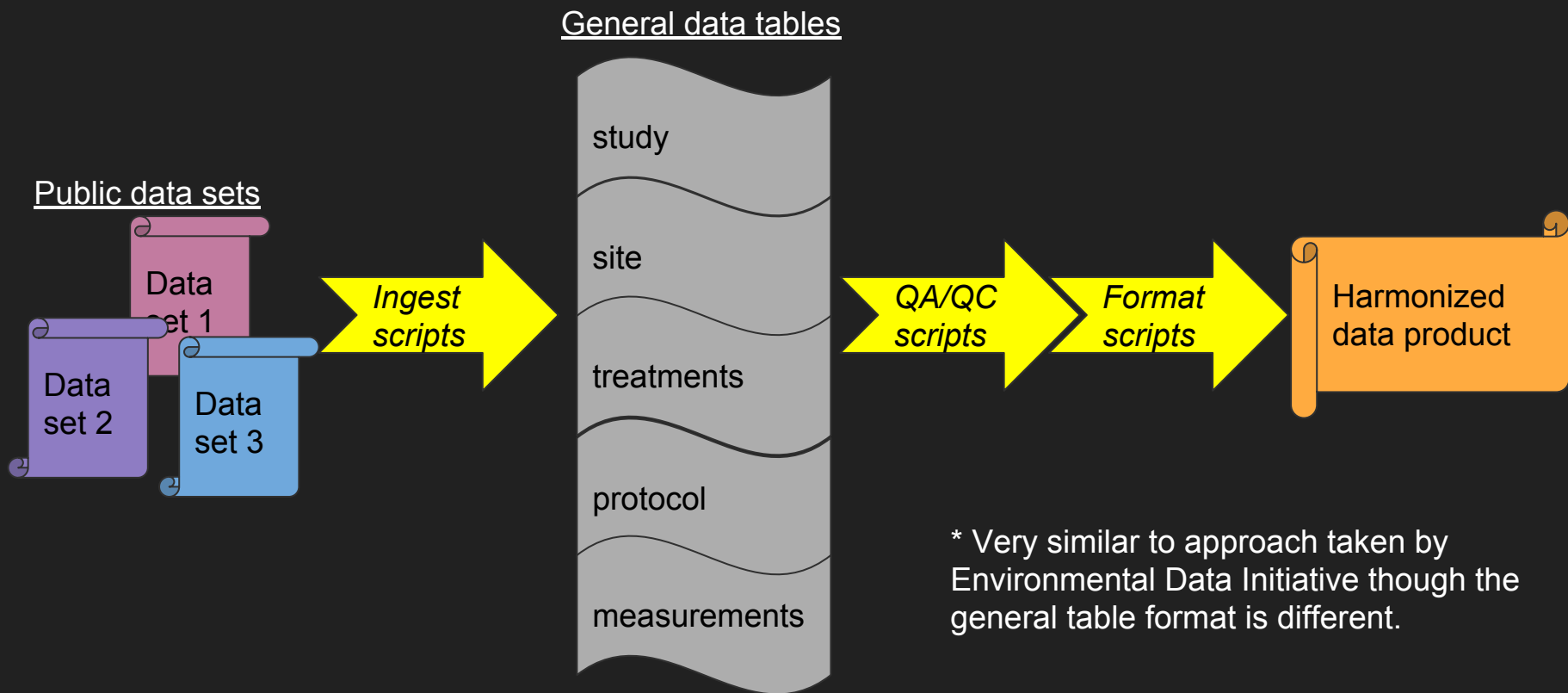
Data providers see...

- 1) Huge list of variables
- 2) What do variables mean and how does this connect to my data?
- 3) Probably can't find a field for some of the data (data loss or scope creep)

Both data providers and data users have responsibilities

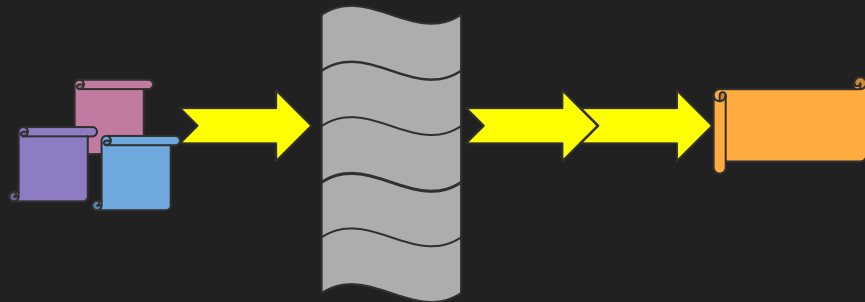


SOC-DRaHR: A new data pipeline



SOC-DRaHR: A new data pipeline

1. Identified datasets of interest
2. Data ingest scripts
3. QA/QC scripts
4. Formatting scripts



Set of scripts and meta-data.

Contributors to any of the 4 steps get a co-author on next DOI.

ISCN will periodically archive data product as permissions allow.

Data script is washing glassware.

The mechanics are not computationally challenging.

Requires domain knowledge.

Excellent opportunity to pick up new skills for domain scientists (R, data management, data visualization).

SOC-DRaHR currently: identify and ingest

Current status

Broad community interest and support

20 datasets identified

4 ingestion scripts written (mostly for
ISCN next generation of data products)

General key-ingest script developed

1st hackathon December 2017

Future direction

Finalize format script for ISCN4

1st DOI release Summer 2018

Identify new potential data products

Develop guides and documentation

2nd hackathon May 2018 at Texas A&M

<https://github.com/ISCN/SOC-DRaHR>

<https://github.com/ISCN/soilDataR>

We welcome contributions

- 1) Id data
- 2) Write scripts
- 3) Develop documentation
- 4) Host hackathons

<https://github.com/ISCN/SOC-DRaHR>

<https://github.com/ISCN/soilDataR>

Kathe Todd-Brown
Pacific Northwest National Lab
@KatheMathBio
katherine.todd-brown@pnnl.gov
ktoddbrown@gmail.com
<https://github.com/ktoddbrown>

Thank you!

ISCN co-chairs:

Gustaf Hugelius (Stockholm U), Jennifer Harden (Stanford/USGS)

ISCN co-coordinators:

Avni Malhotra (ORNL)

Repository and other contributions: Rose Abramoff (LBNL), Ben Bond-Lamberty (PNNL), Julie Loisel (Texas A&M)

